

SCIM-D-19-01397: Response to Initial Critique

George Chacko

March 30, 2020

Dear Dr. Glänzel:

Thank you very much for your message of March 6, 2020 in which it was communicated that the manuscript would benefit from major revisions. We agree with this assessment and thank the reviewers for their constructive critique. Accordingly, I am (also on behalf of my co-authors) resubmitting (i) a revised manuscript with markup (ii) a clean copy with only the revisions. Lastly, I have addressed the comments of the reviewers in the remainder of this document.

I hope you will find that our revisions improve the manuscript. We would have responded sooner but the distractions of COVID-19 manifested themselves in a number of ways that impacted normal function. I also hope that you and your colleagues are safe and healthy.

Sincerely

George Chacko, PhD

Reviewer 1

The authors have extended the works of Salton and Bergmark on the computer science research. This study has potential. However, there are several concerns which reviewer wants to mention as follows:

1. Section 2 is really overwhelming with numerical values. Sometimes, its difficult to follow up the discussion. It would be easy for the reader if authors can make unique way of explaining the cleaning and merging they have done. May be in the form of a table to mention the numerical values and a flow chart to explain the flow of the process till the agglomerative clustering.

The reviewer makes a good point. We rewritten Section 2 (Materials and Methods) to include a paragraph that emphasizes flow while not eliminating any of the details we think are necessary. We think it's more readable now.

2. Section-wise brief discussion is missing at the end of Introduction.

We have revised the Introduction to better track the flow of information through the document

3. Quality of some images is really poor. Please improve them to make it much more visible.

We are a little puzzled by this comment, since the pdfs generated on our machines show acceptable image quality. However, we have regenerated all the images in pdf (and TIFF @600 dpi). In some cases, modified axis labels to make them more readable. Thank you for flagging this defect. In the event that the manuscript is accepted for publication, we will certainly work closely with journal staff to generate even higher resolution figures.

4. Why is "Results" not a separate section?

This has been corrected, thank you. The error was on account of LaTeX formatting.

5. Overall, the paper is interesting covering the field of computer science, which actually needs a separate attention, considering a lot of inter-disciplinary works fading away the actual computer science contribution. However, the presentation of this paper could be simplified for the broader audience.

Thank you for this comment, which is also raised by Reviewer 2. We have broadened the outreach in this revision.

Reviewer 2

This manuscript overview the fields of computer science by the means of clustering for networks of direct citation or co-citation. Its originality is using DBLP bibliography, and I understand that the authors worked hard to process big data. As a result, I judge that this manuscript should be accepted. However, there are some points to be revised. (C01 - C05 are mandatory to be revised.)

Thank you for these comments. We would add respectfully that in terms of originality, our implementation of co-citation based clustering builds on rather than simply reproduces Small and Sweeney's approach.

- C01. The header "Results" has a lack of numbering, and "3 Discussion" will be "4 Discussion".

Thank you for pointing this out. Our apologies, this has been fixed

- C02. Some second-level headers in the sections "2 Materials and Methods" and "Results" don't have period, so add period like "Overview."

Again, our thanks. We have corrected our previously sloppy editing for which we apologize

- C03. Update: "(Fig. 3, Cluster 17), ." – > "(Fig. 3, Cluster 17)."

A correction has been made

- C04. For the section "1 Introduction," what is the purpose of this research, or what is this research aiming at? Ultimately, it is estimated that the part "we extend the work of Salton and Bergmark" will be the purpose. However, readers cannot understand the originality or significance of this research, because it seems to simply apply an existing methodology to current data. Thus, you should describe what you want to know from your results using the work of Salton and Bergmark. First, you should make the purpose specific in the section "1 Introduction". Second, answer to your new purpose in the section "3 Discussion" or others (i.g. "Conclusion" if you add). I recommend that you write an outspoken expression using "purpose" or "aim at" about your purpose for appealing your research's value to roughly-looking readers.

We interpret this comment from the reviewer as noting a lack of communication of purpose in the manuscript. We agree that in trying to reach out to multiple readerships with sometimes thinly overlapping interests, we have presented a broad sense of purpose. Therefore, we have tried to clarify purpose in the revision.

Regarding 'existing methodology to current data.', we will also offer that while we use Graculus, an existing methodology, for clustering by direct citation, it has not been used exhaustively in citation analysis as far as we can tell, and this is consistent

with text in Šubelj, van Eck, and Waltman (2016). Further, we have modified co-citation based clustering as described by Small and Sweeney (1985). Lastly, that clustering methods tend to display sensitivity to input data such that every case has to be carefully evaluated. However, it is our responsibility to communicate this to the reader, so the reviewer’s point is very well taken

- C05. How do the authors identify the field labels? The existing field labels would not be able to used, because your new borders by clustering and those of the existing field are different.

We are sorry that this was not clear. Yes, the new borders are different. Therefore, in each case we we only considered ‘Scopus ASJC minor subject area categories that accounted for at least 15% of the publications in each cluster’. Thus, a cluster could map to multiple subject areas if at least 15% of its content matched to each of them. We had to try setting different thresholds, indeed we show the heatmap for co-citation is shown at two different thresholds.

- C06. [Optional] In the section “3 Discussion,” you had better to write a limitation by excluding small fields, which may include innovative seeds. In order to use this research for the purpose of promoting academic-industry collaboration, you can also describe it as your future work. I like Fig.3 (c), because I’m excited to low score areas that may be new opportunities. Meanwhile, currently, agricultural IT is emerging and fishery starts to use IT. You may be able to provide information of new opportunities with your cutoff small fields by using figures like Fig.3 (c)

For reasons of scope and feasibility, we restricted ourselves to interpreting strong signals from a measurement at a single point in time. Further, ‘To focus on relatively high signal, we only considered Scopus ASJC minor subject area categories that accounted for at least 15% of the publications in each cluster’, but we have added a sentence to the Discussion about excluding small fields, which may (or may not) include innovative seeds

- C07. [Optional] As a significance of your research, the authors can suggest the academic about a question if some current systems that a paper connects to one field are outdated.

Even though we find the idea attractive, we are hesitant to verge into speculation. We do make the point that contextualized interpretation of these results is important and will address the suggestion to a greater extent in the revision

- I’m looking forward to reviewing the updated manuscript again.

Thank you!